SIAM J. MATRIX ANAL. APPL.
Vol. 33, No. 1, pp. 52–72

© 2012 Society for Industrial and Applied Mathematics

DIFFERENCE FILTER PRECONDITIONING FOR LARGE COVARIANCE MATRICES*

MICHAEL L. STEIN†, JIE CHEN‡, AND MIHAI ANITESCU‡

Abstract. In many statistical applications one must solve linear systems involving large, dense, and possibly irregularly structured covariance matrices. These matrices are often ill-conditioned; for example, the condition number increases at least linearly with respect to the size of the matrix when observations of a random process are obtained from a fixed domain. This paper discusses a preconditioning technique based on a differencing approach such that the preconditioned covariance matrix has a bounded condition number independent of the size of the matrix for some important process classes. When used in large scale simulations of random processes, significant improvement is observed for solving these linear systems with an iterative method.

Key words. condition number, preconditioner, stochastic process, random field, spectral analysis, fixed-domain asymptotics

AMS subject classifications. 65F35, 60G25, 62M15

DOI. 10.1137/110834469

1. Introduction.

A problem that arises in many statistical applications is the solution of linear systems of equations for large positive definite covariance matrices (see, e.g., [20]). An underlying challenge for solving such linear systems is that covariance matrices are often dense and ill-conditioned. Specifically, if one considers taking an increasing number of observations of some random process in a fixed and bounded domain, then one often finds the condition number grows without bound at some polynomial rate in the number of observations. This asymptotic approach, in which an increasing number of observations is taken in a fixed region, is called fixed-domain asymptotics. It is used extensively in spatial statistics [20] and is being increasingly used in time series, especially in finance, where high frequency data is now ubiquitous [3]. Preconditioned iterative methods are usually the practical choice to solve for these covariance matrices, whereby the matrix-vector multiplications and the choice of a preconditioner are two crucial factors that affect the computational efficiency. Whereas the former problem has been extensively explored, for example, by using the fast multipole method [12, 4, 8], the latter has not acquired satisfactory answers yet. Some designs of the preconditioners, which are relevant to this work given the close connection between linear systems arising from radial basis function (RBF) interpolation and the covariance matrix, have been proposed and analyzed empirically (see, e.g., [5, 9, 13]); however, their behavior was rarely theoretically studied.

---

[*] Received by the editors May 19, 2011; accepted for publication (in revised form) by J. G. Nagy October 19, 2011; published electronically January 5, 2012. The submitted manuscript has been created by the University of Chicago as Operator of Argonne National Laboratory ("Argonne") under contract DE-AC02-06CH11357 with the U.S. Department of Energy. The U.S. Government retains for itself, and others acting on its behalf, a paid-up, nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the government.

http://www.siam.org/journals/simax/33-1/83446.html

[†] Department of Statistics, University of Chicago, Chicago, IL 60637 (stein@galton.uchicago.edu). The work of this author was supported by U.S. Department of Energy grant DE-SC0002557.

[‡] Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL 60439 (jiechen@mcs.anl.gov, anitescu@mcs.anl.gov). The work of these authors was supported by the U.S. Department of Energy under contract DE-AC02-06CH11357.

Furthermore, the preconditioners used therein are tuned towards specific iterative solvers (e.g., GMRES as in [5], an iterative technique that is analogous to conjugate gradient as in [9, 13]), and they are not symmetric and might lose positive definiteness when the interpolation constraints are removed. Therefore, further analysis and adaptation in design are needed before they can be applied to a covariance matrix in an iterative solver that exploits the symmetric definiteness of a matrix.

This paper proves that for processes whose spectral densities decay at certain specific rates at high frequencies, the preconditioned covariance matrices have a bounded condition number. The preconditioners use filters based on simple differencing operations, which have long been used to "prewhiten" (make the covariance matrix closer to a multiple of the identity) regularly observed time series. However, the utility of such filters for irregularly observed time series and spatial data is not as well recognized. These cases are the focus of this work.

Consider a stationary real-valued random process $Z(x)$ with covariance function $k(x)$ and spectral density $f(\omega)$, which are mutually related by the Fourier transform and the inverse transform:

$$k(x) = \int_{-\infty}^{+\infty} f(\omega) \exp(i\omega x)\, d\omega, \qquad f(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} k(x) \exp(-i\omega x)\, dx.$$

In higher dimensions, the process is more often called a *random field*. Where boldface letters denote vectors, a random field $Z(\boldsymbol{x})$ in $\mathbb{R}^d$ has the following covariance function $k(\boldsymbol{x})$ with spectral density $f(\boldsymbol{\omega})$:

$$k(\boldsymbol{x}) = \int_{\mathbb{R}^d} f(\boldsymbol{\omega}) \exp(i\boldsymbol{\omega}^T \boldsymbol{x})\, d\boldsymbol{\omega}, \qquad f(\boldsymbol{\omega}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} k(\boldsymbol{x}) \exp(-i\boldsymbol{\omega}^T \boldsymbol{x})\, d\boldsymbol{x}.$$

For real-valued processes, both k and f are even functions. This paper describes results for irregularly sited observations in one dimension and for gridded observations in higher dimensions. To facilitate the presentation, we will, in general, use the notation for d dimensions, except when discussions or results are specific to one dimension. The covariance matrix K for observations $\{Z(\boldsymbol{x}_j)\}$ at locations $\{\boldsymbol{x}_j\}$ is defined as

$$K(j,l) \equiv \operatorname{cov}\{Z(\boldsymbol{x}_j), Z(\boldsymbol{x}_l)\} = k(\boldsymbol{x}_j - \boldsymbol{x}_l).$$

Taking f to be nonnegative and integrable guarantees that k is a valid covariance function. Indeed, for any real vector $\boldsymbol{a}$,

$$\boldsymbol{a}^T K \boldsymbol{a} = \sum_{j,l} a_j a_l k(\boldsymbol{x}_j - \boldsymbol{x}_l) = \int_{\mathbb{R}^d} f(\boldsymbol{\omega}) \left| \sum_j a_j \exp(i\boldsymbol{\omega}^T \boldsymbol{x}_j) \right|^2 d\boldsymbol{\omega}, \tag{1.1}$$

which is obviously nonnegative as it must be since it equals $\operatorname{var}\left\{ \sum_j a_j Z(\boldsymbol{x}_j) \right\}$. The existence of a spectral density implies that k is continuous.

In some statistical applications, a family of parameterized covariance functions is chosen, and the task is to estimate the parameters and to uncover the underlying covariance function that presumably generates the given observed data. Let $\boldsymbol{\theta}$ be the vector of parameters. We expand the notation and denote the covariance function by $k(\boldsymbol{x}; \boldsymbol{\theta})$. Similarly, we use $K(\boldsymbol{\theta})$ to denote the covariance matrix parameterized by $\boldsymbol{\theta}$. We assume that observations $y_j = Z(\boldsymbol{x}_j)$ come from a stationary random field that is Gaussian with zero mean.[1] The maximum likelihood estimation [17] method estimates the parameter $\boldsymbol{\theta}$ by finding the maximizer of the log-likelihood function

---

[1] The case of nonzero mean that is linear in a vector of unknown parameters can be handled with little additional effort by using maximum likelihood or restricted maximum likelihood [20].

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{1}{2}\boldsymbol{y}^T K(\boldsymbol{\theta})^{-1}\boldsymbol{y} - \frac{1}{2}\log(\det(K(\boldsymbol{\theta}))) - \frac{m}{2}\log 2\pi,$$

where the vector $\boldsymbol{y}$ contains the m observations $\{y_j\}$. A maximizer $\hat{\boldsymbol{\theta}}$ is called a maximum likelihood estimate of $\boldsymbol{\theta}$. The optimization can be performed by solving (assuming there is a unique solution) the score equation

$$-\boldsymbol{y}^T K(\boldsymbol{\theta})^{-1}\frac{\partial K(\boldsymbol{\theta})}{\partial \theta_\ell}K(\boldsymbol{\theta})^{-1}\boldsymbol{y} + \operatorname{tr}\left(K(\boldsymbol{\theta})^{-1}\frac{\partial K(\boldsymbol{\theta})}{\partial \theta_\ell}\right) = 0 \qquad \forall\,\ell, \tag{1.2}$$

where the left-hand side is nothing but the partial derivative of $-2\mathcal{L}(\boldsymbol{\theta})$. Because of the difficulty of evaluating the trace for a large implicitly defined matrix, Anitescu, Chen, and Wang [1] exploited the Hutchinson estimator[2] of the matrix trace and proposed solving the sample average approximation of the score equation instead:

$$\begin{aligned} F_\ell(\boldsymbol{\theta}) := -\boldsymbol{y}^T K(\boldsymbol{\theta})^{-1}\frac{\partial K(\boldsymbol{\theta})}{\partial \theta_\ell}K(\boldsymbol{\theta})^{-1}\boldsymbol{y} \\ + \frac{1}{N}\sum_{j=1}^{N}\boldsymbol{u}_j^T\left(K(\boldsymbol{\theta})^{-1}\frac{\partial K(\boldsymbol{\theta})}{\partial \theta_\ell}\right)\boldsymbol{u}_j = 0 \qquad \forall\,\ell, \end{aligned} \tag{1.3}$$

where the sample vectors $\boldsymbol{u}_j$'s have independent Rademacher variables as entries. As the number N of sample vectors tends to infinity, the solution $\hat{\boldsymbol{\theta}}^N$ of (1.3) converges to $\hat{\boldsymbol{\theta}}$ in distribution:

$$(V^N/N)^{-1/2}(\hat{\boldsymbol{\theta}}^N - \hat{\boldsymbol{\theta}}) \xrightarrow{\mathcal{D}} \text{standard normal}, \tag{1.4}$$

where V^N is some positive definite matrix dependent on the Jacobian and the variance of $\boldsymbol{F}(\boldsymbol{\theta})$. This error needs to be distinguished from the error in $\hat{\boldsymbol{\theta}}$ itself as an estimate of $\boldsymbol{\theta}$. Roughly speaking, this convergence result indicates that the ℓth estimated parameter $\hat{\theta}_\ell^N$ has variance of approximately $V^N(\ell,\ell)/N$ when N is sufficiently large. Practical approaches (such as a Newton-type method) for solving (1.3) will need to evaluate $\boldsymbol{F}$ (possibly multiple times), which in turn requires solving a linear system involving K with multiple right-hand sides ($\boldsymbol{y}$ and $\boldsymbol{u}_j$'s).

If we do not precondition, the condition number of K grows faster than linearly in m assuming the observation domain has finite diameter and k is continuous, which holds for any integrable spectral density f. To prove this, first note that we can pick observation locations $\boldsymbol{y}_m$ and $\boldsymbol{z}_m$ among $\boldsymbol{x}_1, \dots, \boldsymbol{x}_m$ such that $|\boldsymbol{y}_m - \boldsymbol{z}_m| \to 0$ as $m \to \infty$ and k continuous implies $\operatorname{var}\left\{\frac{1}{\sqrt{2}}Z(\boldsymbol{y}_m) - \frac{1}{\sqrt{2}}Z(\boldsymbol{z}_m)\right\} \to 0$ as $m \to \infty$, so that the minimum eigenvalue of K also tends to 0 as $m \to \infty$. To get a lower bound on the maximum eigenvalue, we note that there exists $r > 0$ such that $k(\boldsymbol{x}) > \frac{1}{2}k(\boldsymbol{0})$ for all $|\boldsymbol{x}| \le r$. Assume that the observation domain has a finite diameter, so that it can be covered by a finite number of balls of diameter r and call this number B. Then for any m, one of these balls must contain at least $m' \ge m/B$ observations. The sum of these observations divided by $\sqrt{m'}$ has a variance of at least $\frac{1}{2}m'k(\boldsymbol{0}) \ge \frac{m}{2B}k(\boldsymbol{0})$, so the maximum eigenvalue of K grows at least linearly with m. Thus, the ratio of the maximum to the minimum eigenvalue of K and hence its condition number grows faster than linearly in m. How much faster clearly depends on the smoothness of Z, but we will not pursue this topic further here.

---

[2] One can also use other stochastic estimators to approximate the matrix trace, which result in different convergence rates; see [2] for an exposition of the estimators. We note that the Hutchinson estimator with $N \approx 100$ yielded a satisfactory approximation to the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ in a problem with m large [1].

In what follows, we consider a filtering technique that essentially preconditions K such that the new system has a condition number that does not grow with the size of K for some distinguished process classes. Strictly speaking, the filtering operation, though linear, is not equal to a preconditioner in the standard sense, since it reduces the size of the matrix by a small number. Thus, we also consider augmenting the filter to obtain a full-rank linear transformation that serves as a real preconditioner. However, as long as the rank of the filtering matrix is close to m, maximum likelihood of $\boldsymbol{\theta}$ based on the filtered observations should generally be nearly as statistically effective as maximum likelihood based on the full data. In particular, maximum likelihood estimates are invariant under full-rank transformations of the data.

The theoretical results on bounded condition numbers heavily rely on the properties of the spectral density f. For example, the results in one dimension require that the process behaves not too differently than does either Brownian motion or integrated Brownian motion, at least at high frequencies. Although the restrictions on f are strong, they do include some models frequently used for continuous time series and in spatial statistics. As noted earlier, the theory is developed based on fixed-domain asymptotics; without loss of generality, we assume that this domain is the box $[0, T]^d$. As the observations become denser, for continuous k the correlations of neighboring observations tend to 1, resulting in matrices K that are nearly singular. However, the proposed difference filters can precondition K so that the resulting matrix has a bounded condition number independent of the number of observations. Section 4 gives several numerical examples demonstrating the effectiveness of this preconditioning approach.

2. Filter for a one-dimensional case.

Let the process $Z(x)$ be observed at locations

$$0 \le x_0 < x_1 < \cdots < x_n \le T,$$

and suppose the spectral density f satisfies

$$f(\omega)\omega^2 \text{ bounded away from 0 and } \infty \text{ as } \omega \to \infty. \tag{2.1}$$

The spectral density of Brownian motion is proportional to ω^{-2}, so (2.1) says that Z is not too different from Brownian motion in terms of its high frequency behavior. Define the process filtered by differencing and scaling as

$$Y_j^{(1)} = [Z(x_j) - Z(x_{j-1})]/\sqrt{d_j}, \qquad j = 1, \ldots, n, \tag{2.2}$$

where $d_j = x_j - x_{j-1}$. Let $K^{(1)}$ denote the covariance matrix of the $Y_j^{(1)}$'s:

$$K^{(1)}(j, l) = \operatorname{cov}\left\{Y_j^{(1)}, Y_l^{(1)}\right\}.$$

For Z Brownian motion, $K^{(1)}$ is a multiple of the identity matrix, and (2.1) is sufficient to show the condition number of $K^{(1)}$ is bounded by a finite value independent of the number of observations.

THEOREM 2.1. *Suppose Z is a stationary process on $\mathbb{R}$ with spectral density f satisfying* (2.1). *There exists a constant C depending only on T and f that bounds the condition number of $K^{(1)}$ for all n.*

If we let $L^{(1)}$ be a bidiagonal matrix with nonzero entries

$$L^{(1)}(j,j-1) = -1/\sqrt{d_j} \qquad \text{and} \qquad L^{(1)}(j,j) = 1/\sqrt{d_j},$$

it is not hard to see that K and $K^{(1)}$ are related by

$$K^{(1)} = L^{(1)} K {L^{(1)}}^T.$$

Note that $L^{(1)}$ is rectangular, since the row index ranges from 1 to n and the column index ranges from 0 to n. It entails a special property that each row sums to zero:

$$\boldsymbol{a}^T L^{(1)} \mathbf{1} = 0 \tag{2.3}$$

for any vector $\boldsymbol{a}$, where $\mathbf{1}$ denotes the vector of all 1's. It will be clear later that (2.3) is key to the proof of the theorem. For now we note that if $\boldsymbol{\lambda} = {L^{(1)}}^T \boldsymbol{a}$, then

$$\boldsymbol{a}^T K^{(1)} \boldsymbol{a} = \boldsymbol{\lambda}^T K \boldsymbol{\lambda} = \operatorname{var}\left\{ \sum_j \lambda_j Z(x_j) \right\} \quad \text{with } \sum_j \lambda_j = 0. \tag{2.4}$$

Strictly speaking, ${L^{(1)}}^T L^{(1)}$ is not a preconditioner, since $L^{(1)}$ has more columns than rows, even though the transformed matrix $K^{(1)}$ has a desirable condition property. A real preconditioner can be obtained by augmenting $L^{(1)}$. To this end, we define, in addition to (2.2),

$$Y_0^{(1)} = Z(x_0), \tag{2.5}$$

and let $\tilde{K}^{(1)}$ denote the covariance matrix of all the $Y_j^{(1)}$'s, including $Y_0^{(1)}$. Then we have

$$\tilde{K}^{(1)} = \tilde{L}^{(1)} K {\tilde{L}^{(1)}}^T,$$

where $\tilde{L}^{(1)}$ is obtained by adding to $L^{(1)}$ the 0th row, with 0th entry equal to 1 and other entries 0. Clearly, $\tilde{L}^{(1)}$ is nonsingular. Thus, ${\tilde{L}^{(1)}}^T \tilde{L}^{(1)}$ preconditions the matrix K.

COROLLARY 2.2. *Suppose Z is a stationary process on $\mathbb{R}$ with spectral density f satisfying* (2.1). *Then there exists a constant C depending only on T and f that bounds the condition number of $\tilde{K}^{(1)}$ for all n.*

For a given situation, it may turn out that $\operatorname{var}(Y_0^{(1)})$ is much different from the variances of the normalized first differences, leading to a covariance matrix with large condition number. In this case, and similarly for the augmentation used in Corollary 2.4, a convenient practical fix is to consider the correlation matrix of the transformed observations rather than the covariance matrix (see section 5).

We next consider the case where the spectral density f satisfies

$$f(\omega)\omega^4 \text{ bounded away from 0 and } \infty \text{ as } \omega \to \infty. \tag{2.6}$$

Integrated Brownian motion, a process whose first derivative is Brownian motion, has spectral density proportional to ω^{-4}. Thus (2.6) says Z behaves somewhat like integrated Brownian motion at high frequencies. In this case, the appropriate preconditioner uses second order differences. Define

$$(2.7)\quad Y_j^{(2)} = \frac{[Z(x_{j+1}) - Z(x_j)]/d_{j+1} - [Z(x_j) - Z(x_{j-1})]/d_j}{2\sqrt{d_{j+1} + d_j}}, \qquad j = 1, \ldots, n-1,$$

and denote by $K^{(2)}$ the covariance matrix of the $Y_j^{(2)}$'s, $j = 1, \ldots, n-1$, namely,

$$K^{(2)}(j, l) = \operatorname{cov}\left\{Y_j^{(2)}, Y_l^{(2)}\right\}.$$

Then for Z integrated Brownian motion, $K^{(2)}$ is a tridiagonal matrix with a bounded condition number (see section 2.3). This result allows us to show the condition number of $K^{(2)}$ is bounded by a finite value independent of n whenever f satisfies (2.6).

THEOREM 2.3. *Suppose Z is a stationary process on $\mathbb{R}$ with spectral density f satisfying* (2.6). *Then there exists a constant C depending only on T and f that bounds the condition number of $K^{(2)}$ for all n.*

If we let $L^{(2)}$ be the tridiagonal matrix with nonzero entries

$$\begin{aligned} L^{(2)}(j, j-1) &= 1/(2d_j\sqrt{d_j + d_{j+1}}), \\ L^{(2)}(j, j+1) &= 1/(2d_{j+1}\sqrt{d_j + d_{j+1}}), \\ L^{(2)}(j, j) &= -L^{(2)}(j, j-1) - L^{(2)}(j, j+1) \end{aligned}$$

for $j = 1, \ldots, n-1$, and let $K^{(2)}$ be the covariance matrix of the $Y_j^{(2)}$'s, then K and $K^{(2)}$ are related by

$$K^{(2)} = L^{(2)} K L^{(2)^T}.$$

Similar to (2.3), the matrix $L^{(2)}$ has a property that for any vector $\boldsymbol{a}$,

$$(2.8)\qquad \begin{aligned} \boldsymbol{a}^T L^{(2)} \boldsymbol{x}^0 &= 0, \\ \boldsymbol{a}^T L^{(2)} \boldsymbol{x}^1 &= 0, \end{aligned}$$

where $\boldsymbol{x}^0 = \mathbf{1}$, the vector of all 1's, and $\boldsymbol{x}^1$ has entries $(\boldsymbol{x}^1)_j = x_j$. In other words, if we let $\boldsymbol{\lambda} = L^{(2)^T}\boldsymbol{a}$, then

$$\boldsymbol{a}^T K^{(2)} \boldsymbol{a} = \boldsymbol{\lambda}^T K \boldsymbol{\lambda} = \operatorname{var}\left\{\sum_j \lambda_j Z(x_j)\right\}, \qquad \text{with } \sum_j \lambda_j = 0 \text{ and } \sum_j \lambda_j x_j = 0.$$

To yield a preconditioner for K in the strict sense, in addition to (2.7), we define

$$Y_0^{(2)} = Z(x_0) + Z(x_n) \qquad \text{and} \qquad Y_n^{(2)} = [Z(x_n) - Z(x_0)]/(x_n - x_0).$$

Accordingly, we augment the matrix $L^{(2)}$ to $\tilde{L}^{(2)}$ with

$$\tilde{L}^{(2)}(0, l) = \begin{cases} 1, & l = 0, \\ 1, & l = n, \\ 0 & \text{otherwise,} \end{cases} \qquad \tilde{L}^{(2)}(n, l) = \begin{cases} -1/(x_n - x_0), & l = 0, \\ 1/(x_n - x_0), & l = n, \\ 0 & \text{otherwise} \end{cases}$$

and use $\tilde{K}^{(2)}$ to denote the covariance matrix of the $Y_j^{(2)}$'s, including $Y_0^{(2)}$ and $Y_n^{(2)}$. Then, we obtain

$$\tilde{K}^{(2)} = \tilde{L}^{(2)} K \tilde{L}^{(2)^T}.$$

One can easily verify that $\tilde{L}^{(2)}$ is nonsingular. Thus, $\tilde{L}^{(2)^T} \tilde{L}^{(2)}$ becomes a preconditioner for K.

COROLLARY 2.4. *Suppose Z is a stationary process on $\mathbb{R}$ with spectral density f satisfying* (2.6). *Then there exists a constant C depending only on T and f that bounds the condition number of $\tilde{K}^{(2)}$ for all n.*

We expect that versions of the theorems and corollaries hold whenever, for some positive integer τ, $f(\omega)\omega^{2\tau}$ is bounded away from 0 and ∞ as $\omega \to \infty$. However, the given proofs rely on detailed calculations on the covariance matrices and do not easily extend to larger τ. Nevertheless, we find it interesting and somewhat surprising that no restriction is needed on the spacing of the observation locations, especially for $\tau = 2$. These results perhaps give some hope that similar results for irregularly spaced observations might hold in more than one dimension.

The rest of this section gives proofs of the above results. The proofs make substantial use of results concerning equivalence of Gaussian measures [14]. In contrast, the results for the high dimension case (presented in section 3) are proved without recourse to equivalence of Gaussian measures.

2.1. Intrinsic random function and equivalence of Gaussian measures.

We first provide some preliminaries. For a random process Z (not necessarily stationary) on $\mathbb{R}$ and a nonnegative integer p, a random variable of the form $\sum_{j=1}^n \lambda_j Z(x_j)$ for which $\sum_{j=1}^n \lambda_j x_j^\ell = 0$ for all nonnegative integers $\ell \le p$ is called an authorized linear combination of order p, or ALC-p [7]. If, for every ALC-p $\sum_{j=1}^n \lambda_j Z(x_j)$, the process $Y(x) = \sum_{j=1}^n \lambda_j Z(x + x_j)$ is stationary, then Z is called an intrinsic random function of order p, or IRF-p [7].

Similar to stationary processes, intrinsic random functions have spectral measures, although they may not be integrable in a neighborhood of the origin. We still use $g(\omega)$ to denote the spectral density with respect to the Lebesgue measure. Corresponding to these spectral measures are what are known as generalized covariance functions. Specifically, for any IRF-p, there exists a generalized covariance function $G(x)$ such that for any ALC-p $\sum_{j=1}^n \lambda_j Z(x_j)$,

$$\operatorname{var}\left\{\sum_{j=1}^n \lambda_j Z(x_j)\right\} = \sum_{j,l=1}^n \lambda_j \lambda_l G(x_j - x_l).$$

Although a generalized covariance function G cannot be written as the Fourier transform of a positive finite measure, it is related to the spectral density g by

$$\sum_{j,l=1}^n \lambda_j \lambda_l G(x_j - x_l) = \int_{-\infty}^{+\infty} g(\omega) \left| \sum_{j=1}^n \lambda_j \exp(i\omega x_j) \right|^2 d\omega$$

for any ALC-p $\sum_{j=1}^n \lambda_j Z(x_j)$.

Brownian motion is an example of an IRF-0, and integrated Brownian motion is an example of an IRF-1. Defining $g_r(\omega) = |\omega|^{-r}$, Brownian motion has a spectral density proportional to g_2 with generalized covariance function $-c|x|$ for some $c > 0$. Note that if one sets $Z(0)=0$, then $\operatorname{cov}\{Z(x),Z(s)\}=\min\{x,s\}$ for $x,s\geq 0$. Integrated Brownian motion has a spectral density proportional to g_4 with generalized covariance function $c|x|^3$ for some $c>0$.

We will need to use some results from Stein [21] on the equivalence of Gaussian measures. Let L_T be the vector space of random variables generated by $Z(x)$ for $x\in[0,T]$ and let $L_{T,p}$ be the subspace of L_T containing all ALC-p's in L_T, so that $L_T\supset L_{T,0}\supset L_{T,1}\supset\cdots$. Let $P_{T,p}(f)$ and $P_T(f)$ be the Gaussian measure for $L_{T,p}$ and L_T, respectively, when Z has mean 0 and spectral density f. For measures P and Q on the same measurable space, write $P\equiv Q$ to indicate that the measures are equivalent (mutually absolutely continuous). Since $L_T\supset L_{T,p}$, for two spectral densities f and g, $P_T(f)\equiv P_T(g)$ implies that $P_{T,p}(f)\equiv P_{T,p}(g)$ for all $p\geq 0$.

2.2. Proof of Theorem 2.1. Let $K(h)$ denote the covariance matrix K associated to a spectral density h, and similarly for $K^{(1)}(h)$, $\tilde{K}^{(1)}(h)$, $K^{(2)}(h)$, and $\tilde{K}^{(2)}(h)$. The main idea of the proof is to put upper and lower bounds on the bilinear form $\boldsymbol{a}^TK^{(1)}(f)\boldsymbol{a}$ for f satisfying (2.1) by constants times $\boldsymbol{a}^TK^{(1)}(g_2)\boldsymbol{a}$. Then since $K^{(1)}(g_2)$ has a condition number 1 independent of n, it immediately follows that $K^{(1)}(f)$ has a bounded condition number, also independent of n.

Let $f_0(\omega)=(1+\omega^2)^{-1}$ and

$$f_R(\omega)=\begin{cases} f(\omega), & |\omega|\leq R,\\ f_0(\omega), & |\omega|>R\end{cases}$$

for some R. By (2.1), there exist R and $0<C_0<C_1<\infty$ such that $C_0f_R(\omega)\leq f(\omega)\leq C_1f_R(\omega)$ for all ω. Then by (1.1) and (2.4), for any real vector $\boldsymbol{a}$,

$$C_0\cdot\boldsymbol{a}^TK^{(1)}(f_R)\boldsymbol{a}\leq\boldsymbol{a}^TK^{(1)}(f)\boldsymbol{a}\leq C_1\cdot\boldsymbol{a}^TK^{(1)}(f_R)\boldsymbol{a}. \tag{2.9}$$

By the definition of f_0, we have $P_{T,0}(f_0)\equiv P_{T,0}(g_2)$ [21, Theorem 1]. Since $f_R=f_0$ for $|\omega|>R$, by Ibragimov and Rozanov [14, Theorem 17 of Chapter III], we have $P_T(f_R)\equiv P_T(f_0)$; thus $P_{T,0}(f_R)\equiv P_{T,0}(f_0)$. Therefore, by the transitivity of equivalence, we obtain that $P_{T,0}(f_R)\equiv P_{T,0}(g_2)$. From basic properties of equivalent Gaussian measures (see [14, equation (2.6) on page 76]), there exist constants $0<C_2<C_3<\infty$ such that for any ALC-0, $\sum_{j=0}^n\lambda_jZ(x_j)$ with $0\leq x_j\leq T$ for all j,

$$C_2\operatorname{var}_{g_2}\left\{\sum_{j=0}^n\lambda_jZ(x_j)\right\}\leq\operatorname{var}_{f_R}\left\{\sum_{j=0}^n\lambda_jZ(x_j)\right\}\leq C_3\operatorname{var}_{g_2}\left\{\sum_{j=0}^n\lambda_jZ(x_j)\right\},$$

where var_f, for example, indicates that variances are computed under the spectral density f. Then by (2.4) we obtain

$$C_2\cdot\boldsymbol{a}^TK^{(1)}(g_2)\boldsymbol{a}\leq\boldsymbol{a}^TK^{(1)}(f_R)\boldsymbol{a}\leq C_3\cdot\boldsymbol{a}^TK^{(1)}(g_2)\boldsymbol{a}. \tag{2.10}$$

Combining (2.9) and (2.10), we have

$$C_0C_2\cdot\boldsymbol{a}^TK^{(1)}(g_2)\boldsymbol{a}\leq\boldsymbol{a}^TK^{(1)}(f)\boldsymbol{a}\leq C_1C_3\cdot\boldsymbol{a}^TK^{(1)}(g_2)\boldsymbol{a},$$

and thus the condition number of $K^{(1)}(f)$ is upper bounded by $C_1C_3/(C_0C_2)$.

2.3. Proof of Theorem 2.3. Following a similar argument as in the preceding proof, the bilinear form $\boldsymbol{a}^T K^{(2)}(f)\boldsymbol{a}$ for f satisfying (2.6) can be upper and lower bounded by constants times $\boldsymbol{a}^T K^{(2)}(g_4)\boldsymbol{a}$. Then it suffices to prove that $K^{(2)}(g_4)$ has a bounded condition number, and thus the theorem holds.

To estimate the condition number of $K^{(2)}(g_4)$, first note the fact that for any two ALC-1's $\sum_j \mu_j Z(x_j)$ and $\sum_j \eta_j Z(x_j)$,

$$\sum_{j,l} \mu_j \eta_l (x_j - x_l)^3 = 0. \tag{2.11}$$

Based on the generalized covariance function of g_4, $c|x|^3$, we have

$$\begin{aligned}
&(j,l)\text{-entry of } K^{(2)}(g_4)\\
&\quad = \operatorname{cov}\left\{Y_j^{(2)}, Y_l^{(2)}\right\}\\
&\quad = \operatorname{cov}\left\{\sum_{j'=-1}^{+1} L^{(2)}(j,j+j')Z(x_{j+j'}), \sum_{l'=-1}^{+1} L^{(2)}(l,l+l')Z(x_{l+l'})\right\}\\
&\quad = c\sum_{j'=-1}^{+1}\sum_{l'=-1}^{+1} L^{(2)}(j,j+j')L^{(2)}(l,l+l')|x_{j+j'} - x_{l+l'}|^3.
\end{aligned}$$

Since for any j, $Y_j^{(2)}$ is ALC-1, by using (2.11) one can calculate that

$$(j,l)\text{-entry of } K^{(2)}(g_4) = \begin{cases} c, & l = j,\\ -cd_{j+1}/(2\sqrt{d_{j+1}+d_j}\sqrt{d_{j+2}+d_{j+1}}), & l = j+1,\\ 0, & |l-j| > 1,\end{cases}$$

which means that $K^{(2)}(g_4)$ is a tridiagonal matrix with a constant diagonal c.

To simplify notation, let $C(j,l)$ denote the (j,l)-entry of $K^{(2)}(g_4)$. We have

$$\begin{aligned}
|C(j-1,j)| + |C(j,j+1)| &= \frac{cd_j}{2\sqrt{d_j+d_{j-1}}\sqrt{d_{j+1}+d_j}} + \frac{cd_{j+1}}{2\sqrt{d_{j+1}+d_j}\sqrt{d_{j+2}+d_{j+1}}}\\
&\le \frac{c\sqrt{d_j}}{2\sqrt{d_{j+1}+d_j}} + \frac{c\sqrt{d_{j+1}}}{2\sqrt{d_{j+1}+d_j}} \le \frac{c}{\sqrt{2}}.
\end{aligned}$$

For any vector $\boldsymbol{a}$,

$$\boldsymbol{a}^T K^{(2)}(g_4)\boldsymbol{a} = \sum_{j,l=1}^{n-1} a_j a_l C(j,l) \ge c\sum_{j=1}^{n-1} a_j^2 - 2\sum_{j=1}^{n-2} |a_j a_{j+1} C(j,j+1)|,$$

but

$$\begin{aligned}
2\sum_{j=1}^{n-2} |a_j a_{j+1} C(j,j+1)| &\le \sum_{j=1}^{n-2} (a_j^2 + a_{j+1}^2)|C(j,j+1)|\\
&\le \sum_{j=1}^{n-1} a_j^2 (|C(j-1,j)| + |C(j,j+1)|)\\
&\le \frac{c}{\sqrt{2}} \sum_{j=1}^{n-1} a_j^2.
\end{aligned}$$

Therefore,

$$\boldsymbol{a}^T K^{(2)}(g_4)\boldsymbol{a} \ge c(1-1/\sqrt{2})\,\|\boldsymbol{a}\|^2 . \tag{2.12}$$

Similarly, we have $\boldsymbol{a}^T K^{(2)}(g_4)\boldsymbol{a} \le c(1+1/\sqrt{2})\,\|\boldsymbol{a}\|^2$. Thus the condition number of $K^{(2)}(g_4)$ is at most $(1+1/\sqrt{2})/(1-1/\sqrt{2}) = 3+2\sqrt{2}$.

2.4. Proof of Corollaries 2.2 and 2.4. The proof of Corollary 2.2 is similar to but simpler than the proof of Corollary 2.4 and is omitted. The main idea of proving Corollary 2.4 is to consider the following covariance function (eq. (3.836.5) in [11] with $n = 4$ shows that B_ϕ is a valid covariance function):

$$B_\phi(x) = \begin{cases} \frac{32}{3}\phi^3 - 4\phi x^2 + |x|^3, & |x| \le 2\phi, \\ \frac{1}{3}(4\phi - |x|)^3, & 2\phi < |x| \le 4\phi, \\ 0, & |x| > 4\phi \end{cases}$$

for $\phi > 0$ and the covariance function $E(x) = 3e^{-|x|}(1+|x|)$. The function B_ϕ has a spectral density $h_\phi(\omega)$ proportional to $\sin(\phi\omega)^4/(\phi\omega)^4$, and E has a spectral density $\theta(\omega) = 6\{\pi(1+\omega^2)\}^{-2}$. Using similar ideas as in the proof of Theorem 2.1, we define

$$\theta_R(\omega) = \begin{cases} f(\omega), & |\omega| \le R, \\ \theta(\omega), & |\omega| > R \end{cases}$$

for some R. Then by (2.6), there exist R and $0 < C_0 < C_1 < \infty$ such that for any real vector $\boldsymbol{a}$,

$$C_0 \cdot \boldsymbol{a}^T \tilde{K}^{(2)}(\theta_R)\boldsymbol{a} \le \boldsymbol{a}^T \tilde{K}^{(2)}(f)\boldsymbol{a} \le C_1 \cdot \boldsymbol{a}^T \tilde{K}^{(2)}(\theta_R)\boldsymbol{a}. \tag{2.13}$$

Furthermore, according to the results in [14, Theorem 17 of Chapter III], when $T \le 2\phi$, $P_T(h_\phi) \equiv P_T(\theta) \equiv P_T(\theta_R)$, which leads to

$$C_2 \cdot \boldsymbol{a}^T \tilde{K}^{(2)}(h_\phi)\boldsymbol{a} \le \boldsymbol{a}^T \tilde{K}^{(2)}(\theta_R)\boldsymbol{a} \le C_3 \cdot \boldsymbol{a}^T \tilde{K}^{(2)}(h_\phi)\boldsymbol{a} \tag{2.14}$$

for some $0 < C_2 < C_3 < \infty$. Combining (2.13) and (2.14), it remains to prove that $\tilde{K}^{(2)}(h_\phi)$ has a bounded condition number, since then so does $\tilde{K}^{(2)}(f)$.

When $T \le 2\phi$, only the branch $|x| \le 2\phi$ of B_ϕ is used, and one can compute the covariance matrix $\tilde{K}^{(2)}(h_\phi)$ according to the definition of B_ϕ entry by entry:

$$\begin{aligned}
(0,0)\text{-entry} &= \frac{128}{3}\phi^3 - 8\phi D^2 + 2D^3, && \\
(0,j)\text{-entry} &= \left(-4\phi + \frac{3}{2}D\right)\sqrt{d_{j+1}+d_j} && \text{for } j = 1,\dots,n-1, \\
(0,n)\text{-entry} &= 0, && \\
(j,0)\text{-entry} &= (0,j)\text{-entry} && \text{for } j = 1,\dots,n-1, \\
(j,l)\text{-entry} &= (j,l)\text{-entry of } K^{(2)}(g_4)/c && \text{for } j,l = 1,\dots,n-1, \\
(j,n)\text{-entry} &= (n,j)\text{-entry} && \text{for } j = 1,\dots,n-1, \\
(n,0)\text{-entry} &= 0, && \\
(n,j)\text{-entry} &= \frac{\sqrt{d_{j+1}+d_j}}{D}\left(x_{j-1} + x_j + x_{j+1} - \frac{3}{2}x_0 - \frac{3}{2}x_n\right) && \text{for } j = 1,\dots,n-1, \\
(n,n)\text{-entry} &= 8\phi - 2D, &&
\end{aligned}$$

where $D = x_n - x_0$, and recall that c is the coefficient in the generalized covariance function corresponding to g_4. To simplify notation, let $H(j, l)$ denote the (j, l)-entry of $\tilde{K}^{(2)}(h_\phi)$. Then we have

$$\begin{aligned} \boldsymbol{a}^T \tilde{K}^{(2)}(h_\phi)\boldsymbol{a} = a_0^2 H(0,0) + a_n^2 H(n,n) + 2a_0 \sum_{j=1}^{n-1} a_j H(0,j) + 2a_n \sum_{j=1}^{n-1} a_j H(n,j) \\ + \tilde{\boldsymbol{a}}^T K^{(2)}(g_4)\tilde{\boldsymbol{a}}/c, \end{aligned} \tag{2.15}$$

where $\tilde{\boldsymbol{a}}$ is the vector $\boldsymbol{a}$ with a_0 and a_n removed. For every $\alpha > 0$, using $|2xy| \le x^2 + y^2$ and the Cauchy–Schwarz inequality, we have

$$\begin{aligned} \left| 2a_0 \sum_{j=1}^{n-1} a_j H(0,j) \right| &\le \alpha^2 a_0^2 + \frac{1}{\alpha^2} \sum_{j=1}^{n-1} a_j^2 \sum_{j=1}^{n-1} H(0,j)^2 \\ &\le \alpha^2 a_0^2 + \frac{1}{2\alpha^2} D\,(8\phi - 3D)^2 \sum_{j=1}^{n-1} a_j^2. \end{aligned} \tag{2.16}$$

Similarly, for every $\beta > 0$, using $\left| x_{j-1} + x_j + x_{j+1} - \frac{3}{2}x_0 - \frac{3}{2}x_n \right| \le 3D$, we have

$$\begin{aligned} \left| 2a_n \sum_{j=1}^{n-1} a_j H(n,j) \right| &\le \beta^2 a_n^2 + \frac{1}{\beta^2} \sum_{j=1}^{n-1} a_j^2 \sum_{j=1}^{n-1} H(n,j)^2 \\ &\le \beta^2 a_n^2 + \frac{18D}{\beta^2} \sum_{j=1}^{n-1} a_j^2. \end{aligned} \tag{2.17}$$

Furthermore, by (2.12),

$$\tilde{\boldsymbol{a}}^T K^{(2)}(g_4)\tilde{\boldsymbol{a}}/c \ge (1 - 1/\sqrt{2})\,\|\tilde{\boldsymbol{a}}\|^2\,. \tag{2.18}$$

Applying (2.16), (2.17), and (2.18) to (2.15), together with $D \le T \le 2\phi$, we obtain

$$\begin{aligned} \boldsymbol{a}^T \tilde{K}^{(2)}(h_\phi)\boldsymbol{a} &\ge \left(\frac{128}{3}\phi^3 - 8\phi D^2 + 2D^3 - \alpha^2 \right) a_0^2 + (8\phi - 2D - \beta^2)a_n^2 \\ &\quad + \left(1 - \frac{1}{\sqrt{2}} - \frac{1}{2\alpha^2} D\,(8\phi - 3D)^2 - \frac{18D}{\beta^2} \right) \sum_{j=1}^{n-1} a_j^2 \\ &\ge \left(\frac{128}{3}\phi^3 - 8\phi T^2 + 2T^3 - \alpha^2 \right) a_0^2 + (8\phi - 2T - \beta^2)a_n^2 \\ &\quad + \left(1 - \frac{1}{\sqrt{2}} - \frac{1}{2\alpha^2} T\,(8\phi - 3T)^2 - \frac{18T}{\beta^2} \right) \sum_{j=1}^{n-1} a_j^2. \end{aligned}$$

Setting $\phi = 14T$, $\alpha^2 = 116000T^3$, and $\beta^2 = 100T$ yields

$$\boldsymbol{a}^T \tilde{K}^{(2)}(h_\phi)\boldsymbol{a} \ge \frac{2902}{3} T^3 a_0^2 + 10T a_1^2 + \left(\frac{178359}{232000} - \frac{1}{\sqrt{2}} \right) \sum_{j=1}^{n-1} a_j^2.$$

Since $\frac{178359}{232000} - \frac{1}{\sqrt{2}} \ge .06$, the minimum eigenvalue of $\tilde{K}^{(2)}(h_\phi)$ is bounded away from 0 independent of n. Similarly, the maximum eigenvalue is bounded from above, and thus $\tilde{K}^{(2)}(h_\phi)$ has a bounded condition number.

3. Filter for d-dimensional case.

The results in section 2 do not easily extend to higher dimensions. One simple exception is when the observation locations are the tensor product of d one-dimensional grids and when the covariance function $k(\boldsymbol{x})$ is the product of d covariance functions for each dimension (i.e., $k(\boldsymbol{x}) = k_1(x_1)k_2(x_2)\dots k_d(x_d)$), in which case the covariance matrix is the Kronecker product of d one-dimensional covariance matrices, each of which can be preconditioned separately. However, such models are of questionable relevance in applications [20]. Here, we restrict the locations of observations to a regular grid. In this case, the second order difference filter becomes the standard discrete Laplace operator. A benefit is that the Laplace operator can be recursively applied many times, resulting in essentially a much higher order difference filtering.

Since the observation locations are evenly spaced, we use $\delta = T/n$ to denote the spacing, where n is the number of observations along each dimension. Thus, the locations are $\{\delta \boldsymbol{j}\}$, $\boldsymbol{0} \le \boldsymbol{j} \le \boldsymbol{n}$. Here, $\boldsymbol{j}$ is a vector of integers, and $\boldsymbol{n}$ means the vector of all n's.[3] Since the locations are on a grid, we use vector indices for convenience. Thus, the covariance matrix K for the observations $\{Z(\delta\boldsymbol{j})\}$ has entries $K(\boldsymbol{j}, \boldsymbol{l}) = k(\delta\boldsymbol{j} - \delta\boldsymbol{l})$. We define the Laplace operator Δ to be

$$\Delta Z(\delta \boldsymbol{j}) = \sum_{p=1}^{d} Z(\delta\boldsymbol{j} - \delta\boldsymbol{e}_p) - 2Z(\delta\boldsymbol{j}) + Z(\delta\boldsymbol{j} + \delta\boldsymbol{e}_p),$$

where $\boldsymbol{e}_p$ denotes the unit vector along the pth coordinate. When the operator is applied τ times, we denote

$$Y_{\boldsymbol{j}}^{[\tau]} = \Delta^{\tau} Z(\delta\boldsymbol{j}).$$

Note that this notation is in parallel to the ones in (2.2) and (2.7), with $[\tau]$ meaning the number of applications of the Laplace operator (instead of the order of the difference), and the index $\boldsymbol{j}$ being a vector (instead of a scalar). In addition, we use $K_d^{[\tau]}$ to denote the covariance matrix of $Y_{\boldsymbol{j}}^{[\tau]}$, $\boldsymbol{\tau} \le \boldsymbol{j} \le \boldsymbol{n} - \boldsymbol{\tau}$:

$$K_d^{[\tau]}(\boldsymbol{j}, \boldsymbol{l}) = \operatorname{cov}\left\{Y_{\boldsymbol{j}}^{[\tau]}, Y_{\boldsymbol{l}}^{[\tau]}\right\},$$

where the subscript d is used to emphasize the dimension of the grid. We have the following result.

THEOREM 3.1. *Suppose Z is a stationary random field on $\mathbb{R}^d$ with spectral density f satisfying*

$$f(\boldsymbol{\omega}) \asymp (1 + \|\boldsymbol{\omega}\|)^{-\alpha}, \tag{3.1}$$

where $\alpha = 4\tau$ for some positive integer τ. Then there exists a constant C depending only on T and f that bounds the condition number of $K_d^{[\tau]}$ for all n.

Recall that for $a(\omega), b(\omega) \ge 0$ for all ω the relationship $a(\omega) \asymp b(\omega)$ indicates that there exist $C_1, C_2 > 0$ such that $C_1 a(\omega) \le b(\omega) \le C_2 a(\omega)$ for all ω.

---

[3] Sometimes, boldface letters denote a vector of the same entries (such as $\boldsymbol{n}$ meaning a vector of all n's). In context, this notation is self-explanatory and not to be confused with the notation of a general vector. Other examples in this paper include $\boldsymbol{1}$ and $\boldsymbol{\tau}$.

It is not hard to verify that $K_d^{[\tau]}$ and K are related by $K_d^{[\tau]} = L_d^{[\tau]} K {L_d^{[\tau]}}^T$, where $L_d^{[\tau]} = L_{n-\tau+1} \cdots L_{n-1} L_n$ and L_s is an $(s-1)^d \times (s+1)^d$ matrix with entries

$$L_s(\boldsymbol{j}, \boldsymbol{l}) = \begin{cases} -2d, & \boldsymbol{l} = \boldsymbol{j}, \\ 1, & \boldsymbol{l} = \boldsymbol{j} \pm \boldsymbol{e}_p, \ p = 1, \ldots, d, \\ 0 & \text{otherwise} \end{cases}$$

for $\mathbf{1} \le \boldsymbol{j} \le \boldsymbol{s} - \mathbf{1}$. One may also want to have a nonsingular $\tilde{L}_d^{[\tau]}$ such that the condition number of $\tilde{L}_d^{[\tau]} K {\tilde{L}_d^{[\tau]}}^T$ is bounded. However, we cannot prove that such an augmentation yields matrices with a bounded condition number, although numerical results in section 5 suggest that such a result may be achievable. Stein [19] applied the iterated Laplacian to gridded observations in d dimensions to improve approximations to the likelihood based on the spatial periodogram and similarly made no effort to recover the information lost by using a less-than-full-rank transformation. It is worth noting that processes with spectral densities of the form (3.1) observed on a grid bear some resemblance to Markov random fields [18], which provide an alternative way to model spatial data observed at discrete locations. Furthermore, recent work [15] indicates that Markov random fields can be used to approximate Gaussian fields with spectral densities of the form (3.1) even when observations are not on a grid, and thus provide a different approach to approximating likelihoods than is presented here.

3.1. Proof of Theorem 3.1.

First note that if one restricts to observations on the grid $\delta \boldsymbol{j}$ for $\boldsymbol{j} \in \mathbb{Z}^d$, the covariance function k can be written as an integral in $[-\pi, \pi]^d$:

$$k(\delta \boldsymbol{j}) = \int_{\mathbb{R}^d} f(\boldsymbol{\omega}) \exp(i \boldsymbol{\omega}^T (\delta \boldsymbol{j}))\, d\boldsymbol{\omega} = \int_{[-\pi,\pi]^d} f_\delta(\boldsymbol{\omega}) \exp(i \boldsymbol{\omega}^T \boldsymbol{j})\, d\boldsymbol{\omega},$$

where

$$f_\delta(\boldsymbol{\omega}) = \delta^{-d} \sum_{\boldsymbol{l} \in \mathbb{Z}^d} f(\delta^{-1}(\boldsymbol{\omega} + 2\pi \boldsymbol{l})). \tag{3.2}$$

Denote by $k^{[\tau]}$ the covariance function such that $k^{[\tau]}(\delta \boldsymbol{j} - \delta \boldsymbol{l}) = K_d^{[\tau]}(\boldsymbol{j}, \boldsymbol{l})$. Then according to the definition of the operator Δ, we have $k^{[0]} = k$ and the recurrence

$$\begin{aligned} k^{[\tau+1]}(\delta \boldsymbol{j}) = \sum_{p,q=1}^{d} & k^{[\tau]}(\delta \boldsymbol{j} + \delta(\boldsymbol{e}_p + \boldsymbol{e}_q)) - 2k^{[\tau]}(\delta \boldsymbol{j} + \delta \boldsymbol{e}_p) + k^{[\tau]}(\delta \boldsymbol{j} + \delta(\boldsymbol{e}_p - \boldsymbol{e}_q)) \\ & - 2k^{[\tau]}(\delta \boldsymbol{j} + \delta \boldsymbol{e}_q) + 4k^{[\tau]}(\delta \boldsymbol{j}) - 2k^{[\tau]}(\delta \boldsymbol{j} - \delta \boldsymbol{e}_q) \\ & + k^{[\tau]}(\delta \boldsymbol{j} + \delta(-\boldsymbol{e}_p + \boldsymbol{e}_q)) - 2k^{[\tau]}(\delta \boldsymbol{j} - \delta \boldsymbol{e}_p) + k^{[\tau]}(\delta \boldsymbol{j} + \delta(-\boldsymbol{e}_p - \boldsymbol{e}_q)). \end{aligned}$$

If we let

$$k^{[\tau]}(\delta \boldsymbol{j}) = \int_{[-\pi,\pi]^d} f_\delta^{[\tau]}(\boldsymbol{\omega}) \exp(i \boldsymbol{\omega}^T \boldsymbol{j})\, d\boldsymbol{\omega},$$

then the above recurrence for $k^{[\tau]}$ translates into

$$f_\delta^{[\tau]}(\boldsymbol{\omega}) = \left[\sum_{p=1}^{d} 4 \sin^2 \left(\frac{\omega_p}{2} \right) \right]^{2\tau} f_\delta(\boldsymbol{\omega}), \tag{3.3}$$

and for any real vector $\boldsymbol{a}$, we have

$$\boldsymbol{a}^T K_d^{[\tau]} \boldsymbol{a} = \sum_{\boldsymbol{\tau} \le \boldsymbol{j}, \boldsymbol{l} \le \boldsymbol{n} - \boldsymbol{\tau}} a_{\boldsymbol{j}} a_{\boldsymbol{l}} k^{[\tau]}(\delta \boldsymbol{j} - \delta \boldsymbol{l}) = \int_{[-\pi,\pi]^d} f_\delta^{[\tau]}(\boldsymbol{\omega}) \left| \sum_{\boldsymbol{\tau} \le \boldsymbol{j} \le \boldsymbol{n} - \boldsymbol{\tau}} a_{\boldsymbol{j}} \exp(i \boldsymbol{\omega}^T \boldsymbol{j}) \right|^2 d\boldsymbol{\omega}.$$

Therefore, to prove that $K_d^{[\tau]}$ has a bounded condition number, we need to bound the expression for $\boldsymbol{a}^T K_d^{[\tau]} \boldsymbol{a}$ given in the above equality.

According to the assumption of f in (3.1), combining (3.2) and (3.3), we have

$$\delta^{d-\alpha} f_\delta^{[\tau]}(\boldsymbol{\omega}) \asymp \left[\sum_{p=1}^d 4 \sin^2 \left(\frac{\omega_p}{2} \right) \right]^{2\tau} \sum_{\boldsymbol{l} \in \mathbb{Z}^d} (\delta + \|\boldsymbol{\omega} + 2\pi \boldsymbol{l}\|)^{-\alpha} =: h_\delta(\boldsymbol{\omega}).$$

Therefore, there exist $0 < C_0 \le C_1 < \infty$ independent of δ and $\boldsymbol{a}$, such that

$$C_0 H_\delta(\boldsymbol{a}) \le \delta^{d-\alpha} \boldsymbol{a}^T K_d^{[\tau]} \boldsymbol{a} \le C_1 H_\delta(\boldsymbol{a}), \tag{3.4}$$

where

$$H_\delta(\boldsymbol{a}) = \int_{[-\pi,\pi]^d} h_\delta(\boldsymbol{\omega}) \left| \sum_{\boldsymbol{\tau} \le \boldsymbol{j} \le \boldsymbol{n} - \boldsymbol{\tau}} a_{\boldsymbol{j}} \exp(i \boldsymbol{\omega}^T \boldsymbol{j}) \right|^2 d\boldsymbol{\omega}.$$

We proceed to bound the function $H_\delta(\boldsymbol{a})$.

For any $\delta \ne 0$, $h_\delta(\boldsymbol{\omega})$ is continuous with $h_\delta(\boldsymbol{0}) = 0$. When $\delta = 0$ and $\alpha = 4\tau$, it can be shown that $h_0(\boldsymbol{\omega})$ is also continuous, but $h_0(\boldsymbol{0}) = 1$. In other words, h_δ converges to h_0 pointwise except at the origin. Since $h_\delta > h_{\delta'}$ when $\delta < \delta'$, we have that h_δ is upper bounded by h_0 for all δ. Moreover, by the continuity of h_0 in $\boldsymbol{\omega} \in [-\pi, \pi]^d$, h_0 has a maximum C_2. Therefore, $h_\delta(\boldsymbol{\omega}) \le C_2$ for all δ and $\boldsymbol{\omega}$, and thus

$$H_\delta(\boldsymbol{a}) \le C_2 \int_{[-\pi,\pi]^d} \left| \sum_{\boldsymbol{\tau} \le \boldsymbol{j} \le \boldsymbol{n} - \boldsymbol{\tau}} a_{\boldsymbol{j}} \exp(i \boldsymbol{\omega}^T \boldsymbol{j}) \right|^2 d\boldsymbol{\omega} = C_2 (2\pi)^d \sum_{\boldsymbol{\tau} \le \boldsymbol{j} \le \boldsymbol{n} - \boldsymbol{\tau}} a_{\boldsymbol{j}}^2. \tag{3.5}$$

Now we need a lower bound for $H_\delta(\boldsymbol{a})$. First, note that when $\boldsymbol{\omega} \in [-\pi, \pi]^d$,

$$h_\delta(\boldsymbol{\omega}) \ge \operatorname{sinc}^2(1/2) \, \|\boldsymbol{\omega}\|^{4\tau} \, (\delta + \|\boldsymbol{\omega}\|)^{-\alpha}.$$

Therefore, for any $0 < \epsilon \le \pi/\delta$,

$$\begin{aligned} H_\delta(\boldsymbol{a}) &\ge \operatorname{sinc}^2(1/2) \int_{[-\pi,\pi]^d} \left(\frac{\|\boldsymbol{\omega}\|}{\delta + \|\boldsymbol{\omega}\|} \right)^\alpha \left| \sum_{\boldsymbol{\tau} \le \boldsymbol{j} \le \boldsymbol{n} - \boldsymbol{\tau}} a_{\boldsymbol{j}} \exp(i \boldsymbol{\omega}^T \boldsymbol{j}) \right|^2 d\boldsymbol{\omega} \\ &\ge \operatorname{sinc}^2(1/2) \int_{[-\pi,\pi]^d \setminus \{\|\boldsymbol{\omega}\| \le \delta\epsilon\}} \left(\frac{\|\boldsymbol{\omega}\|}{\delta + \|\boldsymbol{\omega}\|} \right)^\alpha \left| \sum_{\boldsymbol{\tau} \le \boldsymbol{j} \le \boldsymbol{n} - \boldsymbol{\tau}} a_{\boldsymbol{j}} \exp(i \boldsymbol{\omega}^T \boldsymbol{j}) \right|^2 d\boldsymbol{\omega} \\ &\ge \operatorname{sinc}^2(1/2) \left(\frac{\epsilon}{1 + \epsilon} \right)^\alpha \int_{[-\pi,\pi]^d \setminus \{\|\boldsymbol{\omega}\| \le \delta\epsilon\}} \left| \sum_{\boldsymbol{\tau} \le \boldsymbol{j} \le \boldsymbol{n} - \boldsymbol{\tau}} a_{\boldsymbol{j}} \exp(i \boldsymbol{\omega}^T \boldsymbol{j}) \right|^2 d\boldsymbol{\omega}. \qquad (3.6) \end{aligned}$$

To obtain a lower bound on this last integral, note that

$$\int_{[-\pi,\pi]^d} \left| \sum_{\boldsymbol{\tau} \le \boldsymbol{j} \le \boldsymbol{n} - \boldsymbol{\tau}} a_{\boldsymbol{j}} \exp(i \boldsymbol{\omega}^T \boldsymbol{j}) \right|^2 d\boldsymbol{\omega} = (2\pi)^d \sum_{\boldsymbol{\tau} \le \boldsymbol{j} \le \boldsymbol{n} - \boldsymbol{\tau}} a_{\boldsymbol{j}}^2$$

and

$$\begin{aligned}\int_{\|\boldsymbol{\omega}\|\le\delta\epsilon}\left|\sum_{\boldsymbol{\tau}\le\boldsymbol{j}\le\boldsymbol{n}-\boldsymbol{\tau}} a_{\boldsymbol{j}}\exp(i\boldsymbol{\omega}^T\boldsymbol{j})\right|^2 d\boldsymbol{\omega} &\le \int_{\|\boldsymbol{\omega}\|\le\delta\epsilon}\left(\sum_{\boldsymbol{\tau}\le\boldsymbol{j}\le\boldsymbol{n}-\boldsymbol{\tau}}|a_{\boldsymbol{j}}|\right)^2 d\boldsymbol{\omega}\\ &\le \int_{\|\boldsymbol{\omega}\|\le\delta\epsilon}(n+1-2\tau)^d\sum_{\boldsymbol{\tau}\le\boldsymbol{j}\le\boldsymbol{n}-\boldsymbol{\tau}} a_{\boldsymbol{j}}^2\,d\boldsymbol{\omega}\\ &= (n+1-2\tau)^d(\delta\epsilon)^d V_d\sum_{\boldsymbol{\tau}\le\boldsymbol{j}\le\boldsymbol{n}-\boldsymbol{\tau}} a_{\boldsymbol{j}}^2\\ &\le (T\epsilon)^d V_d\sum_{\boldsymbol{\tau}\le\boldsymbol{j}\le\boldsymbol{n}-\boldsymbol{\tau}} a_{\boldsymbol{j}}^2,\end{aligned}$$

where V_d is the volume of the d-dimensional unit ball, which is always less than 2^d. Applying these results to (3.6),

$$H_\delta(\boldsymbol{a}) \ge \operatorname{sinc}^2(1/2)\left(\frac{\epsilon}{1+\epsilon}\right)^\alpha\left[(2\pi)^d-(T\epsilon)^d V_d\right]\sum_{\boldsymbol{\tau}\le\boldsymbol{j}\le\boldsymbol{n}-\boldsymbol{\tau}} a_{\boldsymbol{j}}^2.$$

Since this bound holds for any $0<\epsilon\le\pi/\delta$, we specifically let $\epsilon = 1/T$. Then

$$H_\delta(\boldsymbol{a}) \ge C_3\sum_{\boldsymbol{\tau}\le\boldsymbol{j}\le\boldsymbol{n}-\boldsymbol{\tau}} a_{\boldsymbol{j}}^2 \tag{3.7}$$

with

$$C_3 = \frac{\operatorname{sinc}^2(1/2)[(2\pi)^d - V_d]}{(1+T)^\alpha},$$

which is independent of δ.

Combining (3.4), (3.5), and (3.7), we have

$$C_0C_3\|\boldsymbol{a}\|^2 \le \delta^{d-\alpha}\boldsymbol{a}^T K_d^{[\tau]}\boldsymbol{a} \le C_1C_2(2\pi)^d\|\boldsymbol{a}\|^2,$$

which means that the condition number of $K_d^{[\tau]}$ is bounded by $(2\pi)^d C_1C_2/(C_0C_3)$.

4. Numerical experiments.

A class of popularly used covariance functions that are flexible in reflecting the local behavior of spatially varying data is the Matérn covariance model [20, 17]:

$$k(\boldsymbol{x}) = \frac{1}{2^{\nu-1}\Gamma(\nu)}\left(\frac{\sqrt{2\nu}\,\|\boldsymbol{x}\|}{\ell}\right)^\nu \mathcal{K}_\nu\left(\frac{\sqrt{2\nu}\,\|\boldsymbol{x}\|}{\ell}\right),$$

where Γ is the Gamma function and $\mathcal{K}_\nu$ is the modified Bessel function of the second kind of order ν. The parameter ν controls the differentiability of the model, and ℓ is a scale parameter. The corresponding spectral density

$$f(\boldsymbol{\omega}) \propto \left(\frac{2\nu}{\ell^2}+\|\boldsymbol{\omega}\|^2\right)^{-(\nu+d/2)},$$

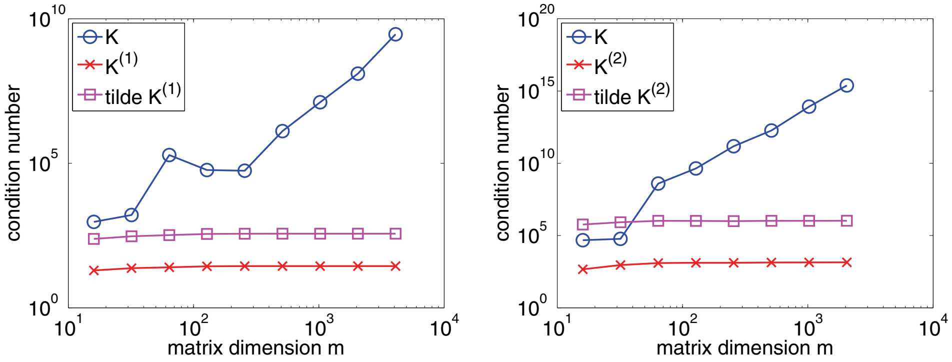
which is dimension dependent. It is clear that with some choices of ν, f satisfies the requirements of the theorems in this paper. For example, when $d=1$, the Matérn model with $\nu = 1/2$ corresponds to Theorem 2.1 and Collorary 2.2, whereas $\nu = 3/2$ corresponds to Theorem 2.3 and Collorary 2.4. Also, when $d = 2$, the Matérn model with $\nu = 1$ corresponds to Theorem 3.1 with $\alpha = 4$, meaning that the Laplace operator Δ is needed to apply once ($\tau = 1$). Whittle [23] argued that the choice of $\nu = 1$ is particularly natural for processes in $\mathbb{R}^2$, in large part because the process is a solution to a stochastic version of the Laplace equation driven by white noise.

Although the restrictions in our theoretical results are quite strong, there are many other models besides Matérn models that satisfy the conditions of the various results. For example, in one dimension, consider rational spectral densities in ω^2, i.e., $f(\omega) = P(\omega^2)/Q(\omega^2)$ for P and Q polynomials of degree p and q, respectively, where P is nonnegative and Q is positive. If $q = p + 1$, then (2.1) is satisfied and if $q = p + 2$, then (2.6) is satisfied. The commonly used spherical covariance functions, given by $k(t) = c\{1 - \frac{3|t|}{2\theta} + \frac{1}{2}\left(\frac{|t|}{\theta}\right)^3\}$ for $|t| < 1$ and 0 otherwise for positive c and θ have spectral densities satisfying (2.1). Another class of covariance functions in one dimension to which it is possible to show (2.1) applies are covariance functions of the form $k(t) = c\left(1 + \frac{|t|}{\theta}\right)^{-\beta}$ for positive c, θ, and β, which [10] shows are positive definite. It is a bit harder to find examples of models used in practice that satisfy (3.1) in more than one dimension, although some rational spectral densities do. In particular, if $f(\boldsymbol{\omega}) = P(\boldsymbol{\omega})/Q(\boldsymbol{\omega})$ for positive even polynomials P and Q with the degree of Q minus the degree of P equaling a multiple of 4, then (3.1) is satisfied. Some non-Matérn models of this form were considered in [22].

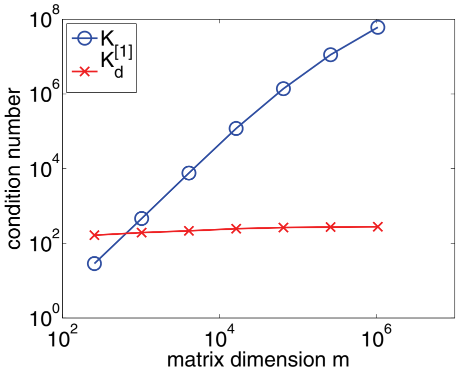
Figure 4.1 plots the curves of the condition numbers for both K and the filtered versions of K, as the size m of the matrix varies for three Matérn models: the first satisfying (2.1), the second (2.6), and the last (3.1) in $d = 2$. The plots were obtained by fixing the domain $T = 100$ and the scale parameter $\ell = 7$. For one-dimensional cases, observation locations were randomly generated according to the uniform distribution on $[0, T]$. The plots clearly show that the condition number of K grows very fast with the size of the matrix. With an appropriate filter applied, on the other hand, the condition number of the filtered covariance matrix stays more or less the same, a phenomenon consistent with the theoretical results.

The good condition property of the filtered covariance matrix is exploited in the block preconditioned conjugate gradient (block PCG) solver. The block version of PCG is used instead of the single vector version because in some applications, such as the one presented in section 1, the linear system has multiple right-hand sides. We remark that the convergence rate of block PCG depends not on the condition number, but on a modified condition number of the linear system [16]. Let λ_j, sorted increasingly, be the eigenvalues of the linear system. With s right-hand sides, the modified condition number is λ_m/λ_s (recall that m is the size of the matrix). Nevertheless, a bounded condition number indicates a bounded modified condition number, which is desirable for block PCG. Figure 4.2 shows the results of an experiment where the observation locations were on a 128×128 regular grid and $s = 100$ random right-hand sides were used. Note that since K and $K_d^{[1]}$ are BTTB (block Toeplitz with Toeplitz blocks), they can be further preconditioned by using a BCCB (block circulant with circulant blocks) preconditioner [6]. Comparing the convergence history for K, K preconditioned with a BCCB preconditioner, $K_d^{[1]}$, and $K_d^{[1]}$ preconditioned with a BCCB preconditioner, we see that the last case clearly yields the fastest convergence.

Next, we demonstrate that the usefulness of the bounded condition number results in the maximum likelihood problem mentioned in section 1. First, observations



(a) $d = 1$, $\nu = 1/2$, first order difference filter. (b) $d = 1$, $\nu = 3/2$, second order difference filter.



(c) $d = 2$, $\nu = 1$, Laplace filter once.

FIG. 4.1. *Condition numbers of K (both unfiltered and filtered) as the matrix size varies.*

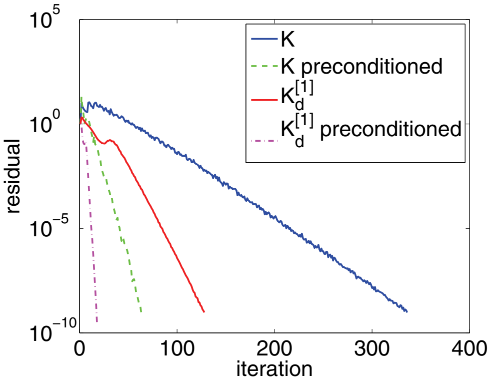


FIG. 4.2. *Convergence history of block PCG.*

$\{y = Z(\boldsymbol{x})\}$ for a Gaussian random field in $\mathbb{R}^2$ are generated under the covariance function

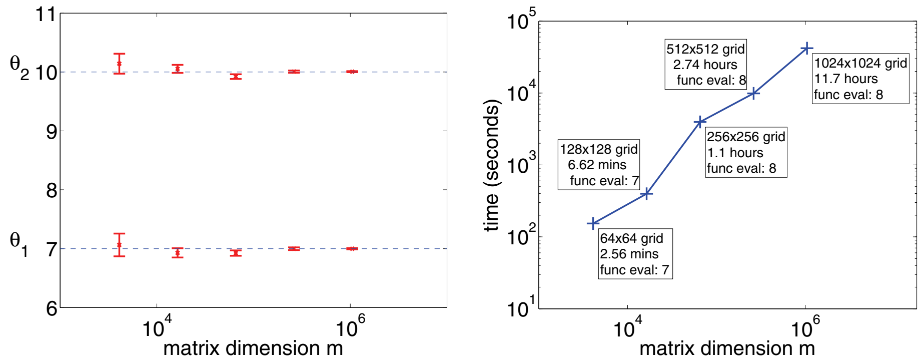
$$k(\boldsymbol{x};\boldsymbol{\theta}) = \sqrt{2} \cdot r_{\boldsymbol{x};\boldsymbol{\theta}} \mathcal{K}_1\left(\sqrt{2} \cdot r_{\boldsymbol{x};\boldsymbol{\theta}}\right), \qquad r_{\boldsymbol{x};\boldsymbol{\theta}} = \sqrt{\frac{x_1^2}{\theta_1^2} + \frac{x_2^2}{\theta_2^2}},$$

where $\boldsymbol{\theta}_* = [7, 10]$ and the observation locations $\boldsymbol{x}$ are on a two-dimensional regular grid of spacing $\delta = 100/n$. To proceed without filtering, one could solve the nonlinear system (1.3). However, as already noted, the condition number of the covariance matrix of y grows faster than linearly with m, so we did not pursue this possibility. We instead solved a nonlinear system other than (1.3) to obtain the estimate $\hat{\boldsymbol{\theta}}^N$. We applied the Laplace operator Δ to the sample vector $\boldsymbol{y}$ once and obtained a vector $\boldsymbol{y}^{[1]}$. Then we solved the nonlinear system

$$-(\boldsymbol{y}^{[1]})^T (K_d^{[1]})^{-1} \frac{\partial K_d^{[1]}}{\partial \theta_\ell} (K_d^{[1]})^{-1} (\boldsymbol{y}^{[1]}) + \frac{1}{N} \sum_{j=1}^{N} \boldsymbol{u}_j^T \left((K_d^{[1]})^{-1} \frac{\partial K_d^{[1]}}{\partial \theta_\ell} \right) \boldsymbol{u}_j = 0, \tag{4.1}$$

where the $\boldsymbol{u}_j$'s are as in (1.3). This approach is equivalent to estimating the parameter $\boldsymbol{\theta}$ from the sample vector $\boldsymbol{y}^{[1]}$ with covariance $K_d^{[1]}$. The matrix $K_d^{[1]}$ is guaranteed to have a bounded condition number for all m according to Theorem 3.1.

The simulation was performed on a Linux desktop with 16 cores with 2.66 GHz frequency and 32 GB of memory. The nonlinear equation (4.1) was solved by using the MATLAB command `fsolve`, which by default used the trust-region dogleg algorithm. Results are shown in Figure 4.3. As we would expect, as the number m of observations increases, the estimates $\hat{\boldsymbol{\theta}}^N$ tend to become closer to $\boldsymbol{\theta}_*$ which generated the simulation data. Furthermore, despite the fact that $N = 100$ is fixed as m increases, the confidence intervals for $\hat{\boldsymbol{\theta}}^N$ become increasingly narrow as m increases, which suggests that it may not be necessary to let N increase with m to ensure that the simulation error $\hat{\boldsymbol{\theta}}^N - \hat{\boldsymbol{\theta}}$ is small compared to the statistical error $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_*$. Finally, as expected, the running time of the simulation scales roughly $O(m)$, which shows promising practicality for running simulations on much larger grids than 1024×1024.



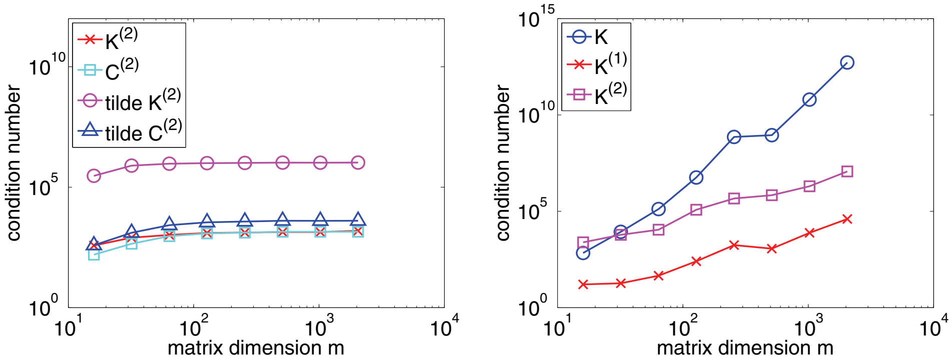
(a) Est. parameters with confidence interval. (b) Running time versus matrix dimension m.

FIG. 4.3. *Simulation results of the maximum likelihood problem.*

5. Further numerical exploration. This section describes additional numerical experiments. First we consider trying to reduce the condition number of our matrices by rescaling them to be correlation matrices. Specifically, for a covariance matrix K, the corresponding correlation matrix is given by

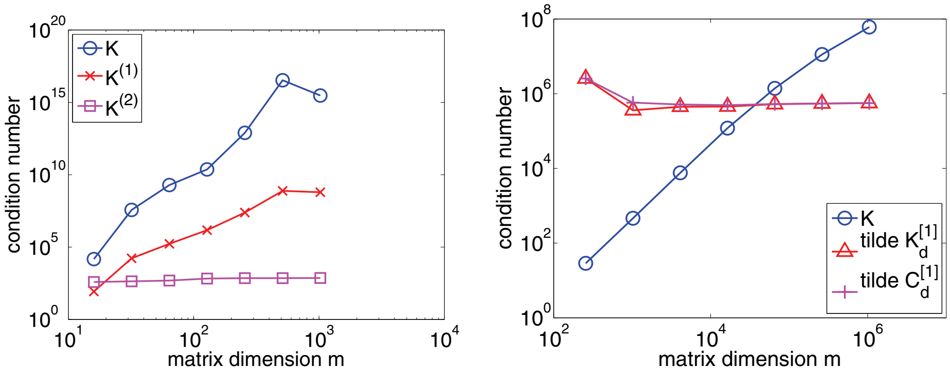
$$C = \operatorname{diag}(K)^{-1/2} \cdot K \cdot \operatorname{diag}(K)^{-1/2}.$$

Although C is not guaranteed to have smaller condition number than K, in practice it often will. For observations on a regular grid and a spatially invariant filter, which is the case in section 3, all diagonal elements of K are equal, so there is no point in rescaling. For irregular observations, rescaling does make a difference. For all of the settings considered in section 2, the ratio of the biggest to the smallest diagonal elements of all of the covariance matrices considered is bounded. It follows that all of the theoretical results in that section on bounded condition numbers apply to the corresponding correlation matrices.



(a) $d = 1$, $\nu = 3/2$, second order difference filter.

(b) $d = 1$, $\nu = 1$, two filters.



(c) $d = 1$, $\nu = 2$, two filters.

(d) $d = 2$, $\nu = 1$, augmented Laplace filter once.

FIG. 5.1. *Condition numbers of covariance matrices and correlation matrices.*

Figure 4.1(b) shows that the filtered covariance matrices $\tilde{K}^{(2)}$ have much larger condition numbers than does $K^{(2)}$. This result is perhaps caused by the full-rank transformation $\tilde{L}^{(2)}$ that makes the $(0,0)$- and (n,n)-entry of $\tilde{K}^{(2)}$ significantly different from the rest of the diagonal. For the same setting, Figure 5.1(a) shows that diagonal rescaling yields much improved results—the correlation matrix $\tilde{C}^{(2)}$ has a condition number much smaller than that of $\tilde{K}^{(2)}$ and close to that of $K^{(2)}$.

Theorems 2.1 and 2.3 indicate the possibility of reducing the condition number of the covariance matrix for spectral densities with a tail similar to $|\omega|^{-p}$ for even p by applying an appropriate difference filter. A natural question is whether the difference filter can also be applied to spectral densities whose tails are similar to $|\omega|$ to some negative odd power. Figures 5.1(b) and 5.1(c) show the filtering results for $|\omega|^{-3}$ and $|\omega|^{-5}$, respectively. In both plots, neither the first nor the second order difference filter resulted in a bounded condition number, but the condition number of the filtered matrix is greatly reduced. This encouraging result indicates that the filtering operation may be useful for a wide range of densities (e.g., all Matérn models) that behave like $|\omega|^{-p}$ at high frequencies, whether or not p is an even integer.

For processes in $d > 1$ dimension, our result (Theorem 3.1) requires a transformation $L_d^{[\tau]}$ that reduces the dimension of the covariance matrix by $O(n^{d-1})$. One may want to have a full-rank transformation or some transformation that reduces the dimension of the matrix by at most $O(1)$. We tested one such transformation here for an $\mathbb{R}^2$ example, which reduced the dimension by four. The transformation $\tilde{L}_d^{[1]}$ is defined as follows. When $\boldsymbol{j}$ is not on the boundary, namely, $\boldsymbol{1} \le \boldsymbol{j} \le \boldsymbol{n} - \boldsymbol{1}$,

$$\tilde{L}_d^{[1]}(\boldsymbol{j}, \boldsymbol{l}) = \begin{cases} -4, & \boldsymbol{l} = \boldsymbol{j}, \\ 2, & \boldsymbol{l} = \boldsymbol{j} + (\pm \boldsymbol{e}_p), \quad p = 1, 2, \\ -1, & \boldsymbol{l} = \boldsymbol{j} + \left[\begin{smallmatrix} \pm 1 \\ \pm 1 \end{smallmatrix} \right], \\ 0 & \text{otherwise.} \end{cases}$$

When $\boldsymbol{j}$ is on the boundary but not at the corner, the definition of $\tilde{L}_d^{[1]}(\boldsymbol{j}, \boldsymbol{l})$ is exactly the same as above, but only for legitimate $\boldsymbol{l}$; that is, components of $\boldsymbol{l}$ cannot be smaller than 0 or larger than n. The corner locations are ignored. The condition numbers of the filtered covariance matrix $\tilde{K}_d^{[1]} = \tilde{L}_d^{[1]} K \tilde{L}_d^{[1]^T}$ and those of the corresponding correlation matrix $\tilde{C}_d^{[1]}$ are plotted in Figure 5.1(d) for the same covariance function used in Figure 4.1(c). Indeed, the diagonal entries of $\tilde{K}_d^{[1]}$ corresponding to the boundary locations are not too different from those not on the boundary; therefore, it is not surprising that the condition numbers for $\tilde{K}_d^{[1]}$ and $\tilde{C}_d^{[1]}$ look similar. It is plausible that the condition number of $\tilde{K}_d^{[1]}$ is bounded independent of the size of the grid.

6. Conclusions.

We have shown that for stationary processes with certain spectral densities, a first/second order difference filter can precondition the covariance matrix of irregularly spaced observations in one dimension, and the discrete Laplace operator (possibly applied more than once) can precondition the covariance matrix of regularly spaced observations in high dimension. Even when the observations are located within a fixed domain, the resulting filtered covariance matrix has a bounded condition number independent of the number of observations. This result is particularly useful for large scale simulations that require the solves of the covariance matrix using an iterative method. It remains to investigate whether the results for high dimension can be generalized for observation locations that are irregularly spaced.

REFERENCES

[1] M. Anitescu, J. Chen, and L. Wang, *A Matrix-Free Approach for Solving the Gaussian Process Maximum Likelihood Problem*, Technical Rep. ANL/MCS-P1857-0311, Argonne National Laboratory, Argonne, IL, 2011.

[2] H. Avron and S. Toledo, *Randomized algorithms for estimating the trace of an implicit symmetric positive semi-definite matrix*, J. ACM, 58 (2011), article 8.

[3] O. E. Barndorff-Nielsen and N. Shephard, *Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics*, Econometrica, 72 (2004), pp. 885–925.

[4] J. Barnes and P. Hut, *A hierarchical $O(N \log N)$ force-calculation algorithm*, Nature, 324 (1986), pp. 446–449.

[5] R. K. Beatson, J. B. Cherrie, and C. T. Mouat, *Fast fitting of radial basis functions: Methods based on preconditioned GMRES iteration*, Adv. Comput. Math., 11 (1999), pp. 253–270.

[6] R. H.-F. Chan and X.-Q. Jin, *An Introduction to Iterative Toeplitz Solvers*, SIAM, Philadelphia, 2007.

[7] J. Chilès and P. Delfiner, *Geostatistics: Modeling Spatial Uncertainty*, John Wiley & Sons, New York, 1999.
[8] Z. Duan and R. Krasny, *An adaptive treecode for computing nonbounded potential energy in classical molecular systems*, J. Comput. Chem., 23 (2001), pp. 1549–1571.
[9] A. C. Faul, G. Goodsell, and M. J. D. Powell, *A Krylov subspace algorithm for multiquadric interpolation in many dimensions*, IMA J. Numer. Anal., 25 (2005), pp. 1–24.
[10] T. Gneiting and M. Schlather, *Stochastic models that separate fractal dimension and the Hurst effect*, SIAM Rev., 46 (2004), pp. 269–282.
[11] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed., Elsevier/ Academic Press, Amsterdam, 2007.
[12] L. Greengard and V. Rokhlin, *A fast algorithm for particle simulations*, J. Comput. Phys., 73 (1987), pp. 325–348.
[13] N. A. Gumerov and R. Duraiswami, *Fast radial basis function interpolation via preconditioned Krylov iteration*, SIAM J. Sci. Comput., 29 (2007), pp. 1876–1899.
[14] I. A. Ibragimov and Y. A. Rozanov, *Gaussian Random Processes*, Springer-Verlag, New York, 1978.
[15] F. Lindgren, H. Rue, and J. Lindström, *An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach*, J. Roy. Statist. Soc. Ser. B, 73 (2011), pp. 423–498.
[16] D. P. O'Leary, *The block conjugate gradient algorithm and related methods*, Linear Algebra Appl., 29 (1980), pp. 293–322.
[17] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, MIT Press, Cambridge, MA, 2006.
[18] H. Rue and L. Held, *Gaussian Markov Random Fields. Theory and Applications*, Chapman & Hall/CRC, Boca Raton, FL, 2005.
[19] M. L. Stein, *Fixed-domain asymptotics for spatial periodograms*, J. Amer. Statist. Assoc., 90 (1995), pp. 1277–1288.
[20] M. L. Stein, *Interpolation of Spatial Data. Some Theory for Kriging*, Springer-Verlag, New York, 1999.
[21] M. L. Stein, *Equivalence of Gaussian measures for some nonstationary random fields*, J. Statist. Plann. Inference, 123 (2004), pp. 1–11.
[22] M. L. Stein, *Space-time covariance functions*, J. Amer. Statist. Assoc., 100 (2005), pp. 310–321.
[23] P. Whittle, *On stationary processes in the plane*, Biometrika, 41 (1954), pp. 434–449.